Харківський національний університет імені В.Н. Каразіна

Факультет математики і інформатики

Кафедра прикладної математики

Кваліфікаційна робота

магістра

на тему «Статистичний аналіз ефективності новітнього препарату проти недрібноклітинного раку легень у другій та третій фазах клінічних досліджень з використанням програмного забезпечення R.»

Виконав:
студентгрупи МП612 курсу магістратури
спеціальність 113Прикладна математика
освітньо-наукова програма
Прикладна математика
Тюрдьо І.М.
Науковий керівник:
професор, доктор фіз.-мат. наук Фардигола Л.В.,
Юсько А.М.
Рецензент:
Principal Statistical Programmer/Analyst Хоба А.М.

Харків - 2024 рік

Анотація

У процесі дослідження ефективності новітнього медичного препарату для лікування недрібноклітинного раку легень буде використана інформація про групу пацієнтів кількістю 224 людини. Цих пацієнтів було за допомогою рандомізації поділено на дві групи , які приймали контрольний, вже існуючий на ринку препарат, та досліджуваний новітній препарат.

Для оцінки ефективності новітнього медичного препарату буде використано метод аналізу виживання. Будуть побудовані криві Каплан-Маєра для функції виживання, сукупної небезпеки, кумулятивних подій, криві виживання по декільком факторам, також буде зроблено лог-ранг тест.

За результатами проведеного дослідження буде зроблено висновок об ефективності досліджуваного препарату в порівнянні з існуючим.

Усі розрахунки та побудови графіків буде зроблено за допомогою програмного забезпечення мови R. Наразі основною програмною мовою при проведенні клінічних медичних досліджень є мова SAS, програмна мова R буде розглянута, як її альтернатива.

In the process of studying the effectiveness of a new drug for the treatment of non-small cell lung cancer, information about a group of patients with 224 people will be collected. These patients were randomized into two groups: they received a control drug already on the market and a follow-up new drug.

To assess the effectiveness of a new medication, a survival analysis method will be used. Kaplan-Maier curves will be generated for the survival function, cumulative risk, cumulative effects, survival curves for several factors, and the log-rank test will also be generated.

The results of the investigation will be followed by a report on the effectiveness of the drug being monitored in combination with the original one.

All routines and daily schedules will be completed using the R software. Although the main software for clinical medical research is SAS, the R software will be considered as an alternative.

Зміст

Вступ

Рак є одним із найбільш серйозних захворювань, з яким стикаються медичні спільноти по всьому світу. Незважаючи на значні покращення в методах діагностики та лікування, рак, як і раніше, залишається однією з головних причин смерті. Розвиток нових лікарських препаратів та методів терапії відіграє ключову роль у боротьбі з цим захворюванням та збільшенні виживання пацієнтів.

Метою даного дослідження є оцінка ефективності нового лікарського препарату антиракової терапії у пацієнтів із недрібноклітинним раком легенів. Для досягнення цієї мети ми використовуємо метод аналізу виживання для оцінки виживання у пацієнтів, які отримують новий препарат, порівняно з пацієнтами, які отримують стандартне лікування. Метод аналізу виживання наразі є одним з головних методів, які використовуються під час проведення клінічних досліджень антиракових препаратів. Його перевагою є здатність враховувати час до настання події (наприклад, смерті від раку), а також на можливість обліку цензурованих даних.

Будуть побудовані криві Каплан-Маєра для функції виживання, сукупної небезпеки, кумулятивних подій, криві виживання по декільком факторам, також буде зроблено лог-ранг тест і модель Кокса, для оцінки виживання та виявлення факторів, що впливають на виживання.

При проведенні дослідження буде використане програмне забезпечення мови R, яка є альтернативою більш уживаній в клінічних дослідженнях мові SAS.

За результатами проведеного дослідження буде зроблено висновок щодо ефективності досліджуваного препарату в порівнянні з існуючим.

Розділ 1. Методи діагностики та лікування недрібноклітинного раку легень

Рак легень підрозділяють на дві основні категорії – дрібноклітинний (близько 20% випадків) і недрібноклітинний (80%).

Недрібноклітинний (NSCLC) тип розвивається при множинних змінах ДНК епітеліальних клітин, внаслідок чого запускається безконтрольний клітинний поділ, втрачається здатність клітини до апоптозу [1].

Виділяють такі різновиди цього раку:

- Аденокарцинома (adenocarcinoma) - цей підтип NSCLC розвивається в епітеліальних клітинах, які виробляють слиз та розташовані в аденоїдних структурах. Він частіше спостерігається у некурців та може бути діагностований за допомогою морфологічних ознак, таких як аденоматозні структури та муцинозні клітини. У порівнянні з іншими видами неклітинного раку поширюється повільно. На цей різновид припадає близько половини випадків недрібноклітинного типу.
- Плоскоклітинний рак (squamous cell carcinoma) - цей підтип NSCLC розвивається в епітеліальних клітинах, які виробляють кератин. Він часто пов'язаний з курінням та може бути діагностований за допомогою морфологічних ознак, таких як плоскі клітини, що утворюють зроговілі пластинки. У харкотинні накопичуються канцерогенні речовини, що призводять до виникнення в клітинах плоского епітелію генетичних мутацій. Локалізований у дихальних шляхах, трапляється у 20-25% випадків. Відрізняється повільним розвитком та безсимптомним перебігом ранніх стадій, що ускладнює своєчасну діагностику.
- Великоклітинний рак (large cell carcinoma) - цей підтип NSCLC характеризується великими анапластичними клітинами з неправильною формою та розмірами. Він часто важко діагностується за допомогою гістологічних ознак та може бути класифікований як великоклітинний аденокарцинома або великоклітинний плоскоклітинний карцинома.

Зустрічається в 10-15% випадків, розвивається в будь-якій ділянці легені. Швидко поширюється, що пояснює складнощі у лікуванні.

- Змішаний (залізисто-плоскоклітинний) рак.

Класифікація стадій проводиться за міжнародною системою TNM, заснованої на розмірах новоутворення, наявності метастазів у лімфатичні вузли та віддалених метастазів, що вражають різні внутрішні органи[2].

Виділяють такі стадії:

- Стадія 0 (карцинома in situ): Пухлина на початковому етапі, коли клітини пухлини не проникли в оточуючі тканини. Вона ще не перешла в стан раку, але може стати раком, якщо не лікувати.
- Стадія I: Рак обмежений в легені та не розповсюджується на сусідні органи чи лімфатичні вузли.
- Стадія II: Рак поширився в легеневі тканини та може впливати на сусідні органи, але ще не досяг регіональних лімфатичних вузлів.
- Стадія III: Рак поширився до регіональних лімфатичних вузлів та може впливати на сусідні структури.

Одна з небезпек цього онкопатології легень у тому, що у перших стадіях захворювання протікає безсимптомно чи викликає слабовиражені хворобливі симптоми, схожі з проявами респіраторних хвороб. Перші симптоми раку легень часто починають турбувати пацієнта, коли злоякісний процес досяг розвиненої стадії.

Недрібноклітинний рак легень (NSCLC) має різноманітні ознаки, які можуть варіювати в залежності від підтипу та стадії захворювання. Ось основні загальні ознаки, які можуть вказувати на наявність NSCLC [3]:

1. Кашель: Постійний кашель, який може бути сухим або з видаленням мокроти. Кашель може бути настільки інтенсивним, що спричиняє біль або дискомфорт.
2. Кров'яниста мокрота: Кров'яниста або коричнева мокрота може бути ознакою ураження легень раком. Це може бути позитивною ознакою, особливо якщо кров'янисті виділення не пов'язані з іншими патологіями.

3. Схуднення та втрата апетиту: Необґрунтована втрата ваги та втрата апетиту можуть бути ознаками, що є частими у людей з раком.
4. Задишка: Задишка або ускладнення дихання можуть виникати через обмеження простору у легенях або ураження легеневих путів.
5. Біль у грудях: Біль у грудях, який може бути різним за характером та інтенсивністю, може виникати внаслідок тиску, який викликає пухлина.
6. Лихорадка — виникає під час застоювання мокротиння внаслідок звуження бронхіального просвіту, що призводить до інфікування.
7. Втома та слабкість: Люди з раком легень часто відчувають велику втомленість та слабкість, навіть при невеликому фізичному навантаженні.

На пізніх стадіях недрібноклітинний рак легень (NSCLC) може супроводжуватися рядом ускладнень та побічних ефектів, оскільки пухлина може розповсюджуватися на інші органи та системи [4]. Ось деякі можливі ускладнення та супутні стани на пізніх стадіях NSCLC:

1. Метастази: Найбільш поширеним ускладненням на пізніх стадіях є метастазування пухлини на інші органи та тканини, такі як печінка, кістки, мозок, нирки, а також на вузли лімфатичної системи. Це може призводити до різноманітних симптомів та ускладнень, залежно від місця метастазування.
2. Паранеопластичні синдроми: Деякі пацієнти можуть розвивати паранеопластичні синдроми, які викликаються продукцією пухлини певних речовин, що впливають на роботу інших органів та систем. Наприклад, синдром Cushing (збільшення рівня кортизолу), гіперкальцемія (збільшення рівня кальцію) та інші.
3. Пневмонія та інші інфекції: Пневмонія та інші інфекції легень можуть виникати внаслідок зменшення функції легень та зниження імунітету, що може бути спричинено раком та/або лікуванням раку.

4. Дистрофія та втрата маси м'язів: Недостатня споживання їжі та слабкість, спричинені раком та/або лікуванням раку, можуть призводити до дистрофії та втрати маси м'язів.
5. Біль та дискомфорт: Рак може викликати біль та дискомфорт, особливо при наявності метастазів у кістках або інших органах.
6. Порушення дихальної функції: Зростаюча пухлина може призводити до порушень дихальної функції.

Однією з головних причин розвитку патології вважається куріння: близько 80% пацієнтів, у яких діагностовано цей тип раку, є курцями з багаторічним стажем. Особливого ризику схильні ті, хто почав курити ще в дитинстві або в підлітковий період, люди, які щодня викурюють більше пачки сигарет, і ті, хто віддає перевагу сигарет без фільтра або дешевим сортам тютюнових виробів. До групи ризику входять пасивні курці.

Існує кілька факторів, які можуть сприяти розвитку недрібноклітинного раку легень [5].

- Куріння: Куріння є основним фактором ризику для розвитку раку легень. Тютюнопаління містить токсичні речовини, які можуть пошкодити клітини легень та сприяти розвитку раку.
- Пасивне куріння: Вдихання диму від тютюну, який виділяється палінням, також може підвищити ризик розвитку раку легень, навіть у тих, хто не курить.
- Забруднення повітря: Дихання забрудненого повітря, такого як вихлопні гази, хімічні речовини та інші токсичні речовини, може також збільшити ризик розвитку раку легень.
- Генетичні фактори: Деякі генетичні аномалії можуть збільшити схильність до розвитку раку легень.
- Важкий або забруднений повітря на робочому місці: Особи, які працюють у промислових секторах з великою кількістю забруднення повітря, таких як шахти або заводи, можуть бути в більшому ризику.

- Попередні захворювання легень: Історія захворювання легень, таких як хронічний обструктивний захворювання легень (ХОЗЛ), може підвищити ризик розвитку раку легень.
- Впливання радону: Вдихання газоподібного радону, який утворюється при розпаді радіоактивного урану, може бути фактором ризику.
- Інші хвороби та умови: Деякі хвороби та умови, такі як аспірація диму, туберкульоз або інші хронічні захворювання легень, можуть підвищити ризик розвитку раку легень.

Для виявлення патології застосовуються такі діагностичні процедури: [6]

- огляд пацієнта, збирання докладного анамнезу, з'ясування можливого стажу куріння, наявність онкопатологій у близьких родичів;
- рентгенологічне обстеження або комп'ютерна томографія органів грудної клітки;
- бронхоскопія, біопсія з подальшим патогістологічним дослідженням отриманого матеріалу;
- торакоскопія та торакотомія;
- цитологія мокротиння;
- загальний та біохімічний аналіз крові, виявлення в крові онкомаркерів;
- молекулярно-генетичні дослідження злоякісних клітин, виявлення генних мутацій для подальшого підбору ефективних таргетних препаратів.

Варіанти лікування недрібноклітиного раку легенів:

- Хірургічне втручання
- Хіміотерапія – застосування протипухлинних препаратів для знищення ракових клітин. Хіміотерапію можна використовувати окремо або разом з іншими методами лікування.
- Таргетна терапія – нові препарати, що блокують сигнали, які стимулюють ріст ракових клітин.

- Імунотерапія – вид лікування, призначений для підвищення природного захисту організму для боротьби з раком.
- Променева терапія – використання виміряних доз радіації для пошкодження ракових клітин та зупинення їх росту.

Розділ 2. Дослідження ефективності новітнього препарату проти недрібноклітинного раку легень

2.1 Аналіз виживання

В сучасному світі зростає значимість вимог до обґрунтованості ефективності методів лікування різних захворювань. Тому при виборі конкретного методу терапії важливо розглядати не лише його успішність у досягненні ремісії, а й вплив лікування на довгостроковий прогноз захворювання, включаючи загальну виживаність пацієнтів та виживаність від конкретного захворювання.

Аналіз виживання - це метод статистичного аналізу, який розроблено для дослідження, оцінки та порівняння часу, що пройшов до настання певної події (наприклад, поява метастазів, одужання, смерті або загострення захворювання)[7].

Виживання S (T) (Survive) – ймовірність «прожити» час більше t з моменту початку спостереження.

Термін вперше був введений страховими агентами, які оцінювали тривалість життя.

$$S(t) = \frac{число\ переживших\ момент\ t}{загальне\ число\ спостережень} \quad (1)$$

Таким чином, виживання – це ймовірність ненастання події до деякого моменту часу [8].

Властивості функції*S(t):*

S (t)=1, якщо t = 0: на самому початку дослідження очікувана подія не відбулася ні в кого з спостережуваних. Імовірність "дожиття" до цього моменту дорівнює 1.

S(t)= 0, якщо t=∞: в кінці дослідження подія відбулася у всіх спостережуваних. Імовірність "дожиття" до цього моменту дорівнює 0.

Графік функції S(t) – крива виживання відображає ймовірність пережити будь-який з моментів часу t (рис.1). Час може вимірюватися в будь-яких відомих одиницях (дні, місяці та ін).

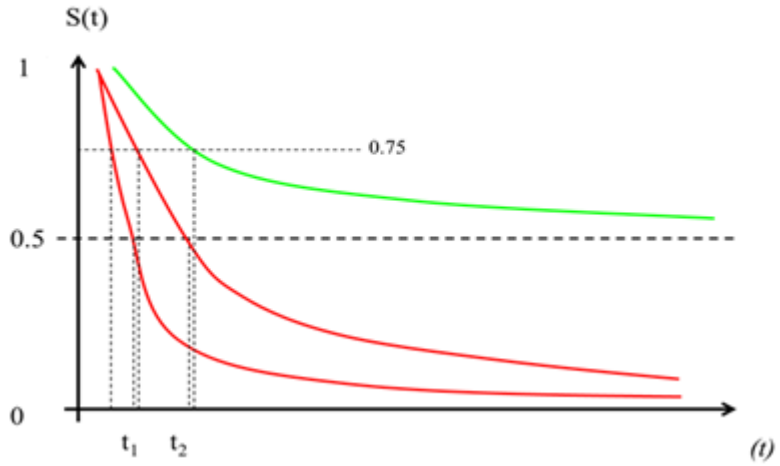

Рисунок1. Крива виживання

Графік виживання може мати різний нахил: крутий графік вказує на низьку виживаність, коли очікувана подія настає швидко у більшості випробовуваних. На відміну, пологий графік вказує на високу виживаність, коли потрібно багато часу, щоб очікувана подія настала у всіх випробовуваних.

Крива виживання використовується для визначення медіани виживання та інших процентилів часу життя.

Медіана виживання - це час, до якого доживає половина випробовуваних. Якщо подія не настала у половини випробовуваних, медіана виживання не визначається. Тоді оцінюється час, до якого "дожили" три чверті всіх випробовуваних (75%). Порівнюючи дві чи більше кривих, за допомогою медіани можна оцінити виживаність у різних групах.

Для побудови кривої виживання використовують метод Каплана-Майєра, коли спостереження починаються та закінчуються в різний час.

Наприклад, пацієнти можуть вийти з лікування перед очікуваним терміном, або ми можемо втратити зв'язок з деякими учасниками дослідження. У таких випадках ми маємо цензуровані дані, де не всі події спостереження відомі. Метод Каплана-Майєра дозволяє ефективно врахувати цю неповноту даних та побудувати криву виживання на основі наявних спостережень, включаючи цензуровані.

Попередньо будується так звана таблиця часу життя.

Таблиця1. Таблица часу життя

Момент часу	Кількість спостережуваних об’єктів до моменту часу*t*	Кількістьподій, які відбулисьв момент часу*t*	Частка тих, хто не досягли події в момент *t*	Виживання (кумулятивна доля)
t	n_i	d_i	$1-\frac{d_i}{n_i}$	*S(t)*

Виживання розраховується як добуток за всіма моментами часу, коли відбулася хоча б одна подія

$$S(t) = \prod(1-\frac{d_i}{n_i}) \quad (2)$$

2.2 Методи оцінки виживання

1. Метод Каплан-Мейєра**.**

Для людей з метою порівняння ефективності та безпеки лікування проводяться контрольовані експерименти, які називаються клінічними випробуваннями[9]. У клінічних або громадських випробуваннях вплив втручання оцінюється шляхом вимірювання кількості випробуваних, що вижили після цього втручання протягом певного періоду часу. Іноді цікаво порівняти виживання випробуваних у двох або більше втручаннях. У ситуаціях, коли виживання є проблемою, то змінна, що цікавить, буде тривалістю часу, який мине до того, як відбудеться якась подія. У багатьох ситуаціях цей проміжок часу дуже довгий, наприклад, в терапії раку. У такому випадку в продовж одиниці часу можна оцінити кількість подій, таких як наприклад смерть. В інших ситуаціях можна оцінити тривалість до рецидиву раку або доки не виникне інфекція. Час, що починається від визначеної точки до виникнення даної події, називається часом виживання [10] та аналіз групових даних як аналіз виживання [11].

Аналіз та моделювання даних про «час до події» є основною метою аналізу виживання. Подія, яку аналізують, може бути різна, наприклад, зникнення пухлини, час виписки з медичного закладу/лікарні, реакція на лікування, смерть або розвиток захворювання. Травма, одужання від хвороби та настання хвороби також можуть бути подією.

Методика аналізу виживання використовується для дослідження тривалості часу до настання певної події, яка є об'єктом інтересу. Вона дозволяє порівнювати цей час між різними групами пацієнтів або оцінювати зв'язок між певними факторами та тривалістю виживання.

Аналіз виживання враховує час, який пройшов до виникнення певної події, і досліджує ймовірність, що ця подія станеться протягом певного періоду.Час виживання – це дані, які вимірюють час до певної події, такої як смерть, невдача, реакція, рецидив або розвиток певного захворювання [12]. Час виживання включає в

себе різні показники, такі як тривалість ремісії, час до зникнення пухлини, час до смерті, а також час від початку лікування до відповіді організму на терапію. Дані про виживання містять інформацію про ці показники, а також характеристики пацієнтів, які впливають на їх виживання та реакцію на лікування. Ці дані можна зібрати з клінічних або епідеміологічних досліджень осіб, які стикаються з гострими або хронічними захворюваннями. В аналізі виживання, на відміну від інших статистичних методів, враховуються як час, так і цензуровані дані, що дозволяє докладніше вивчити тривалість виживання та фактори, що на неї впливають.

Цензуровані дані – це дані, які виникають, коли відомо, що тривалість життя людини відбувається лише протягом певного періоду часу. Існують різні види цензури: права, коли учасник залишається живим до певної дати, ліва, коли подія відбулася до початку дослідження, або інтервальна, коли відомо, що подія відбулася в межах певного інтервалу. Перевагою цензурованих даних є те, що тривалість спостереження може відрізнятися для кожного учасника, і це може бути враховано під час аналізу.

Аналіз виживання може проводитися з фіксованою початковою точкою, де час до виникнення певної події реєструється для учасників. Зазвичай дослідження завершується, коли ще не всі учасники проявляють цю подію, і результат для решти учасників стає невідомим. Також результати тих учасників, які вилучилися з дослідження, залишаються невідомими. Час спостереження записується (цензуровані дані для цих випадків), і отримані дані можна проаналізувати за допомогою методу Каплана-Майєра.

Каплан-Майєр (KM) – це непараметрична оцінка функції виживання, яка зазвичай використовується для опису виживання досліджуваної популяції та порівняння двох досліджуваних популяцій. Оцінка Каплана-Мейєра є потужним статистичним інструментом для аналізу виживання пацієнтів після лікування. Вона дозволяє оцінити ймовірність виживання пацієнтів протягом певного періоду після втручання. У клінічних або громадських дослідженнях ефект втручання оцінюється шляхом вимірювання кількості учасників, які

вижили після цього втручання протягом певного періоду часу. Криві використовуються в оцінці Каплана-Мейєра для визначення подій, цензури та ймовірності виживання.

Крива виживання за Капланом-Майєром використовується в епідеміології для аналізу часу до виникнення певної події та порівняння двох або більше груп суб'єктів. Вона дозволяє визначити частку учасників, які залишаються живими (або не пережили певну подію, наприклад смерть) протягом певного часового періоду.

Оцінка ліміту продукту (PLI) – інша назва оцінки Каплана-Мейєра. Формула ліміту продукту оцінює частку організмів або фізичних пристроїв, які вижили після будь-якого віку, навіть якщо деякі елементи не загинули або вийшли з ладу, а вибірка досить мала [13]. Вона передбачає обчислення послідовних ймовірностей події відбуття протягом певного часу. Ці ймовірності помножуються на раніше обчислені ймовірності для отримання остаточної оцінки. Наприклад, якщо ми розглядаємо виживання жінки з недостатньою фертильністю після лапароскопії та гідротубації, ми можемо визначити ймовірність виживання протягом першого місяця, помноживши ймовірність виживання протягом другого та третього місяців на ймовірність виживання в перший місяць, враховуючи, що жінка вже пережила перші два місяці.

Крива виживання Каплана-Майєра визначається як ймовірність вижити протягом заданого проміжку часу, враховуючи час у багатьох малих інтервалах [11]. У цьому аналізі ми враховуємо три припущення. По-перше, ми вважаємо, що пацієнти, які мають цензуровані дані, мають такі ж перспективи виживання, як і ті, за якими продовжують спостерігати. По-друге, ми припускаємо, що ймовірності виживання однакові для осіб, які включені до дослідження в будь-який момент. По-третє, ми припускаємо, що подія відбувається у вказаний час. Це створює деякі труднощі в тих випадках, коли подія може бути виявлена під час звичайного огляду. Для отримання більш точних прогнозів виживання ми можемо здійснювати більш часті

спостереження за особами протягом коротших проміжків часу, наприклад, щоденно.

2. Метод Кокса (Cox regression) - це статистичний метод, який використовується для аналізу даних виживання та встановлення зв'язку між різними факторами та ймовірністю виживання. Він ґрунтується на припущенні про лінійну залежність між логарифмом відносини шансів та значенням пояснюючих змінних. Основне припущення методу Кокса - пропорційність ризиків. Це означає, що вплив факторів на ймовірність виживання вважається постійним у часі, тобто не змінюється з часом спостереження.

Модель регресії Кокса, відома також як модель пропорційних ризиків (Cox proportional hazards model), досліджує, як час виживання (survival time) залежить від незалежних змінних (predictor variables). Цей напівпараметричний метод передбачає прогнозування ризику настання події (hazard risk) для об'єкта, що розглядається, і оцінює вплив незалежних змінних на цей ризик. При цьому ризик настання події є залежною від часу функцією і виявляє ймовірність настання події для об'єктів, що знаходяться в групі ризику. Ніяких припущень про вид функції інтенсивності/ризику немає, у цьому полягає непараметрична частина способу. Однак усі змінні повинні лінійно впливати на логарифм функції ризику настання події, що становить параметричну компоненту методу [14].

Також у модель включені незалежні змінні (predictors) – характеристики об'єкта (наприклад, вік, стать пацієнта, супутні захворювання та ін.), які можуть впливати на ризик настання події [15].

Потенційно вплив предикторів на виживаність можна оцінити, наприклад, методом лінійної логістичної регресії, запропонованим Девідом Коксом (David Roxbee Cox) в 1958 р. Однак цей метод не допускає наявності неповних спостережень. Напівпараметрична регресійна модель пропорційних ризиків Кокса, запропонована ним у 1972 р., розглядає «небезпеку» (hazard) як міру ризику як функцію, що залежить від часу. Вона моделює вплив предикторів на рівень небезпеки (Hazard rate) з урахуванням періоду

спостереження і допускає наявність неповних (цензурованих) спостережень. Функція ризику h (t) описує миттєву ймовірність події для суб'єктів, які все ще піддаються ризику. Оскільки в даному випадку ймовірність віднесена до часу (інтервалу Δt), h(t) можна трактувати як окремий випадок функції інтенсивності. У загальному вигляді рівняння моделі являє собою добуток двох функцій: базової функції ризику, однакової для всіх спостережень і залежної тільки від часу, експоненти суми досліджуваних предикторів моделі з відповідними коефіцієнтами. Отже, в цьому випадку рівняння є мультиплікативним, що означає, що вплив предикторів множиться на базову функцію ризику. Одним з основних припущень моделі є умова пропорційності ризиків, що означає, що відношення ризику між будь-якими двома об'єктами залишається сталим упродовж часу. Це свідчить про те, що ризик не змінюється зі зміною часу і є незалежним від нього. Згідно з Д. Коксом, якщо припущення про пропорційність ризиків справедливе, можна оцінити вплив предикторів без врахування базової функції ризику. Це означає, що можна дослідити розмір ефекту кожного предиктора незалежно від часу. Таким чином, відсутність зміни ризику в часі є ключовим фактором для використання цього методу. Якщо це припущення про пропорційність ризиків не дотримується, можливе застосування регресії Кокса з коваріатами, залежними від часу. Модель пропорційних ризиків Кокса є напівпараметричною, оскільки не передбачає наявності апріорної інформації про базову функцію відмов (непараметричний компонент), проте визначено вид (логлінійний) регресійної функції (параметричний компонент). Якщо ж визначені обидва компоненти, то модель можна охарактеризувати як параметричну (при цьому її часто називають не моделлю Кокса, а відповідно до розподілу функції небезпеки: наприклад, пропорційних ризиків Вейбулла, Гомперца та ін.). На відміну від оцінки Каплана – Мейєра, регресія Кокса моделює функцію ризику, а не виживання.

Цей підхід дозволяє оцінити відносний ризик виживання між двома або більше групами, але не надає абсолютних значень ризику. Натомість

використовується показник «hazard ratio» (HR), що дозволяє порівнювати темпи виживання між групами. Таким чином, HR вказує на те, наскільки ймовірніше відбувається подія (наприклад, смерть) у одній групі порівняно з іншою). При такому аналізі (Каплана – Мейєра, Нельсона – Аалена та регресійної моделі Кокса) дослідник враховує лише одну «подію інтересу». У випадку, коли відбувається інша (конкуруюча) подія, спостереження припиняються і більше не враховуються в аналізі. Це означає, що події, які спричиняють цензуру, повинні бути незалежними від інтересуючого результату (неінформативна цензура). Тобто, пацієнти, які піддаються цензуруванню в певний момент часу, повинні бути схильні до ризику настання події так само, як і ті пацієнти, які продовжують бути спостереженими. Проте, у реальній практиці така ситуація рідко відбувається. Хворі, які перебувають у списку очікування на трансплантацію, можуть бути виключені з аналізу через загострення супутніх захворювань або отримання трансплантата нирки. Наприклад, смерть пацієнта може виключити можливість трансплантації (як подію інтересу). Коли виникнення інтересуючої події виключається іншою подією, припущення про незалежність порушується (інформативна цензура). Такі події не є повністю незалежними і називаються конкуруючими, а ризик їх настання - конкуруючим ризиком.

Однофакторна регресія Коксу, також відома як модель Коксу пропорційних ризиків, є статистичною моделлю, яка використовується в аналізі виживання для оцінки впливу одного предиктора на виживання пацієнтів. У цій моделі передбачається, що вплив предиктора на ризик події залишається незмінним з часом [18].

Формально модель однофакторної регресії Коксу виглядає наступним чином:

$$h(t|X) = h_0(t) * e^{\beta x} \quad (3)$$

де,

- h(t|X) - інтенсивність події (або ризик події) у момент часу t для пацієнтів із спостережуваним значенням предиктораX,

- $h_0(t)$- базова інтенсивність події, яка може залежати лише від часу,
- β - коефіцієнт регресії, що оцінюється в ході аналізу,
- X – значення предиктора.

Інтерпретація коефіцієнта β у моделі однофакторної регресії Коксу аналогічна інтерпретації коефіцієнтів у лінійній регресії: він показує зміну у логарифмі відношення ризиків (HR) при зміні предиктора на одну одиницю. Якщо β>0, то збільшення значення предиктора збільшує ризик події, а якщо β<0, збільшення значення предиктора зменшує ризик події.

Іншими словами, коефіцієнт ризику вище 1 вказує на коваріату, яка позитивно пов'язана з імовірністю події, і, отже, негативно пов'язана з тривалістю виживання. Підсумовуючи:

HR = 1: немає ефекту

HR < 1: Зменшення небезпеки

HR > 1: Збільшення небезпеки

Для оцінки коефіцієнта β у моделі однофакторної регресії Коксу часто використовується метод часткової правдоподібності. Цей метод дозволяє оцінити коефіцієнти з урахуванням цензурованих даних, які найчастіше зустрічаються в аналізі виживання.

Багатофакторний регресійний аналіз Кокса, також відомий як мультифакторна модель Кокса, є статистичною моделлю, що використовується в аналізі виживання для вивчення впливу кількох незалежних змінних (факторів) на ризик виникнення події (наприклад, смерть) з часом.

Формально, багатофакторна модель Кокса виглядає наступним чином:

$$h(t|X) = h_0(t) * e^{\beta_1 x_1} + e^{\beta_2 x_2} + \cdots + e^{\beta_k x_k} \quad (5)$$

де,

- h(t|X) - інтенсивність події (або ризик події) у момент часу t для пацієнтів зі значеннями предикторів X,
- $h_0(t)$- базова інтенсивність події, яка може залежати тільки від часу,

- X1,X2,...,Xk - значення незалежних змінних (факторів),
- β1,β2,...,βk - координати регресії, які оцінюються під час аналізу.

Кожен βi вказує на те, як змінюється логарифм відношення ризиків (HR) з кожним одиничним збільшенням відповідної незалежної змінної. Якщо βi>0, то збільшення значення фактора призводить до збільшення ризику події, а якщо βi<0, то зменшення значення фактора призводить до зменшення ризику події.

Для оцінки коефіцієнтів β у багатофакторній моделі Кокса також використовується метод часткової правдоподібності. Цей метод дозволяє оцінити коефіцієнти з урахуванням цензурованих даних та різноманітних факторів, що впливають на виживання.

Розділ 3. Програмна реалізація дослідження ефективності новітнього препарату проти недрібноклітинного раку легень

3.1 Принципи застосування програмного забезпечення R для проведення статистичного аналізу

Основні функції аналізу виживання знаходяться в пакетах “survival” та “survminer”.

Основні функції, які використовуються з цих пакетів , включають:

- Surv(): створює об'єкт виживання.
- survfit(): підганяє криву виживання, використовуючи або формулу, або з раніше апроксимованої моделі Кокса.
- coxph(): відповідає моделі регресії пропорційних небезпек Кокса.

Інші додаткові функції, які ви можете використовувати, включають:

- cox.zph(): Перевіряє припущення про пропорційні небезпеки регресійної моделі Кокса.
- survdiff(): Тести на відмінності у виживаності між двома групами з використанням логарифмічного рангу / тесту Мантеля-Генцеля.

Surv() створює змінну відповіді, а типове використання займає час для події, і незалежно від того, чи відбулася подія (тобто смерть проти цензури). survfit() створює криву виживання, яку потім можна відобразити або побудувати графік. coxph() реалізує регресійний аналіз і моделі, визначені так само, як і в звичайних лінійних моделях, але з використанням функції coxph().

3.2 Розрахунок основних показників статистичного аналізу новітнього препарату проти недрібноклітинного раку легенів

Для проведення дослідження ми маємо базу даних з дослідження ефективності застосування різної хіміотерапії при недрібноклітинному раку легень. В дослідженні приймало участь 228 пацієнтів, яких було розподілено за допомогою рандомізації на два різних комплекси хіміотерапії (в подальшому ARM A та ARM B).

Таблиця 2. База даних з дослідження ефективності застосування різної хіміотерапії при недрібноклітинному раку легень.

	siteid	arm	age	sex	race	ph_ecog	calor	wt_loss	cnsr	time
1	3	2	74	1	white	1	1175	NA	2	306
2	3	2	68	2	white	0	1225	15	2	455
3	3	1	56	1	black or african american	0	NA	15	1	1010
4	5	2	57	2	asian	1	1150	11	2	210
5	1	1	60	1	white	0	NA	0	2	883
6	12	1	68	1	asian	1	513	0	1	1022
7	7	1	74	1	black or african american	2	384	10	2	310
8	11	1	71	1	asian	2	538	1	2	361
9	1	1	63	2	asian	1	825	16	2	218
10	7	1	61	2	white	2	271	34	2	166
11	6	2	57	1	black or african american	1	1025	27	2	170
12	16	2	68	1	black or african american	2	NA	23	2	654
13	11	2	68	1	white	1	NA	5	2	728
14	21	1	60	1	black or african american	1	1225	32	2	71
15	12	2	57	1	asian	1	2600	60	2	567

В цій таблиці використовуються наступні змінні:

- siteid - код медичної установи.
- arm – код медичного препарата ARM A = 1, ARM B =2 .
- age – вік у роках.
- race – раса пацієнта.
- ph_ecog – оцінка ефективності за ECOG
- calor – кількість калорій спожитих під час їжі

- wt_loss – втрата ваги
- cnsr - цензура
- time – час до події

3.2.1 Криві виживання

Функція Surv. Це основна функція, яка використовується для створення об'єкта виживання (рис.2).

```
> sfit <- survfit(Surv(time, cnsr) ~ arm, data = lung_n)
> sfit
Call: survfit(formula = Surv(time, cnsr) ~ arm, data = lung_n)

        n events median 0.95LCL 0.95UCL
arm=1 120     93    285     222     348
arm=2 108     72    371     306     477
```

Рисунок 2. Функція Surv.

З допомогою функції summary можна отримати інформацію про таблицю життя.

Таблиця 3.

```
> # Access to the sort summary table
> summary(sfit)$table
      records n.max n.start events    rmean se(rmean) median 0.95LCL 0.95UCL
arm=1     120   120     120     93 331.6822  25.10523    285     222     348
arm=2     108   108     108     72 426.0773  30.04808    371     306     477
```

У цих таблицях відображається рядок для кожного моменту часу, коли відбулася подія або зразок піддавався цензурі. Він показує число в зоні ризику (число все ще залишається) і кумулятивну виживаність на цей момент.

Функції summary() можна надати параметр часу, якщо потрібно відобразити результати пошуку. Також можна створити послідовність чисел, що переходять від одного числа до іншого, з кроком ще одного числа за допомогою функції seq()

```
# ?summary.survfit
range(lung$time)
seq(0, 1100, 100)
summary(sfit, times=seq(0, 1000, 100))
```

Рисунок 3. Функція seq()

Можна використовувати цей вектор послідовності з узагальненим викликом на sfit, щоб отримати таблиці життя в цих інтервалах окремо як для ARM A (1), так і ARM B (2) (рис.4). З цих таблиць можна побачити, що пацієнти, які приймали препарат A, як правило, гірше виживають, ніж ті, які приймали препарат B.

```
> # Summary of survival curves
> summary(sfit)
Call: survfit(formula = Surv(time, cnsr) ~ arm, data = lung_n)

                arm=1
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    5    120       1   0.9917  0.0083       0.9755        1.000
   11    119       1   0.9833  0.0117       0.9607        1.000
   12    118       1   0.9750  0.0143       0.9475        1.000
   13    117       1   0.9667  0.0164       0.9351        0.999
   15    116       1   0.9583  0.0182       0.9232        0.995
   26    115       1   0.9500  0.0199       0.9118        0.990
   30    114       1   0.9417  0.0214       0.9007        0.985
   31    113       1   0.9333  0.0228       0.8898        0.979
   53    112       2   0.9167  0.0252       0.8685        0.967
   59    110       1   0.9083  0.0263       0.8581        0.961
   60    109       1   0.9000  0.0274       0.8479        0.955
                arm=2
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  11    108       2   0.9815  0.0130       0.9564        1.000
  13    106       1   0.9722  0.0158       0.9417        1.000
  54    105       1   0.9630  0.0182       0.9280        0.999
  60    104       1   0.9537  0.0202       0.9149        0.994
  61    103       1   0.9444  0.0220       0.9022        0.989
  65    102       1   0.9352  0.0237       0.8899        0.983
  81    101       1   0.9259  0.0252       0.8778        0.977
```

Рисунок 4. Таблиці життя.

В цій таблиці використовуються наступні змінні:

- time - часові точки.
- n.risk – кількість пацієнтів в зоні ризику в момент t.

- n.event – кількість подій в момент t.
- survival – функція виживання
- std.err - значення середньоквадратичної похибки
- lower 95% CI – нижня границя довірчого інтервалу
- upper 95% CI - верхня границя довірчого інтервалу
-

3.2.2 Криві Каплана-Маєра

Тепер, коли крива виживання підігнана до даних, її досить легко візуалізувати за допомогою графіка Каплана-Мейєра (рис.5).

```
> sfit <- survfit(Surv(time, cnsr)~arm, data=lung_n)
> plot(sfit)
```

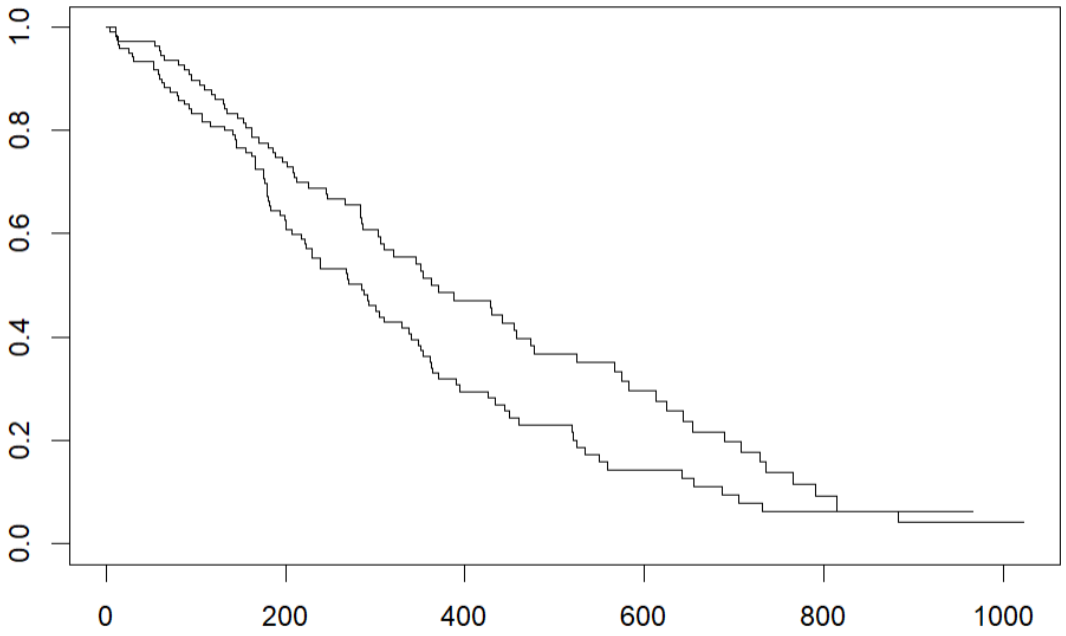

Рисунок 5. Стандартний графік Каплан-Майера.

Існує багато способів модифікації графіка, створеного базовою функцією plot().

Приклад створення графіка з допомогою пакета **survminer**, який надає функцію під назвою **ggsurvplot()** яка значно полегшує створення готових до публікації графіків виживання (рис.6)

```
install.packages("survminer")
library(survminer)
ggsurvplot(sfit)
```

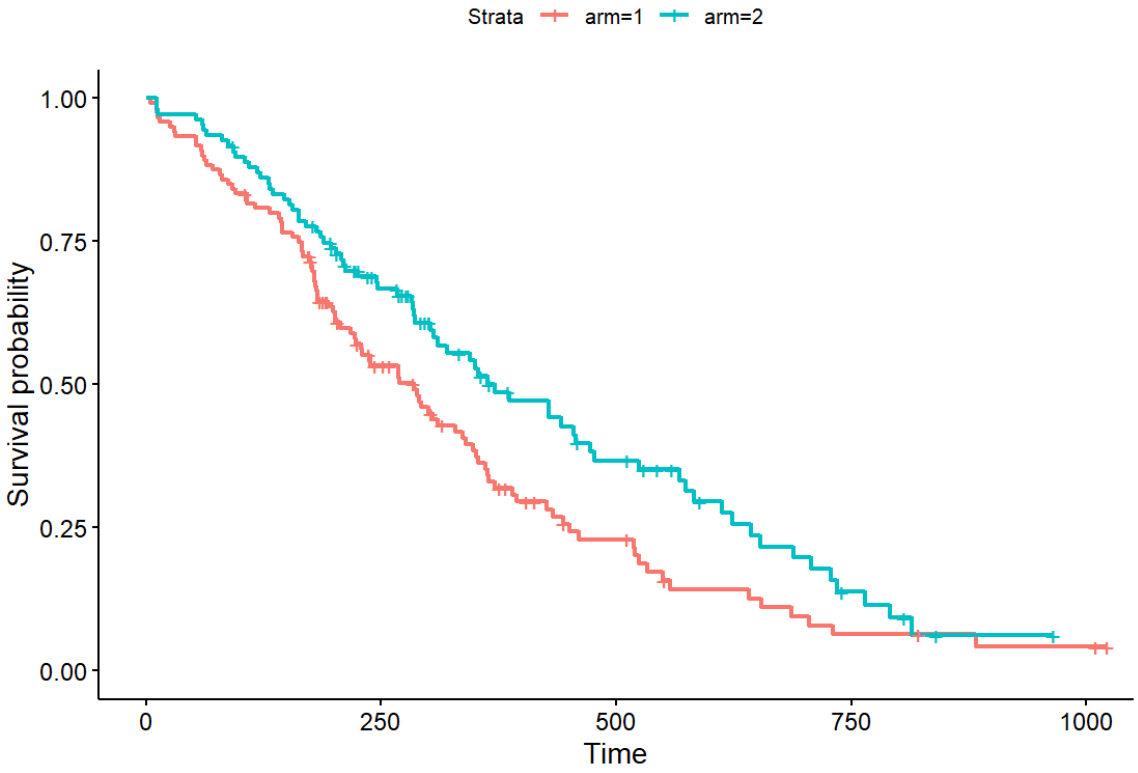

Рисунок 6. Використання **ggsurvplot.**

Цей графік є значно інформативнішим за замовчуванням, просто тому, що він автоматично кодує різні групи кольорами, додає підписи осей та створює та автоматично створює легенду. Але тут можна зробити набагато більше. Можна додати довірчі інтервали, показати p-значення для тесту log-rank, вивести під графіком таблицю ризиків та пацієнтів, котрі були підвергнути цензурі, змінити кольори та мітки груп (рис.6).

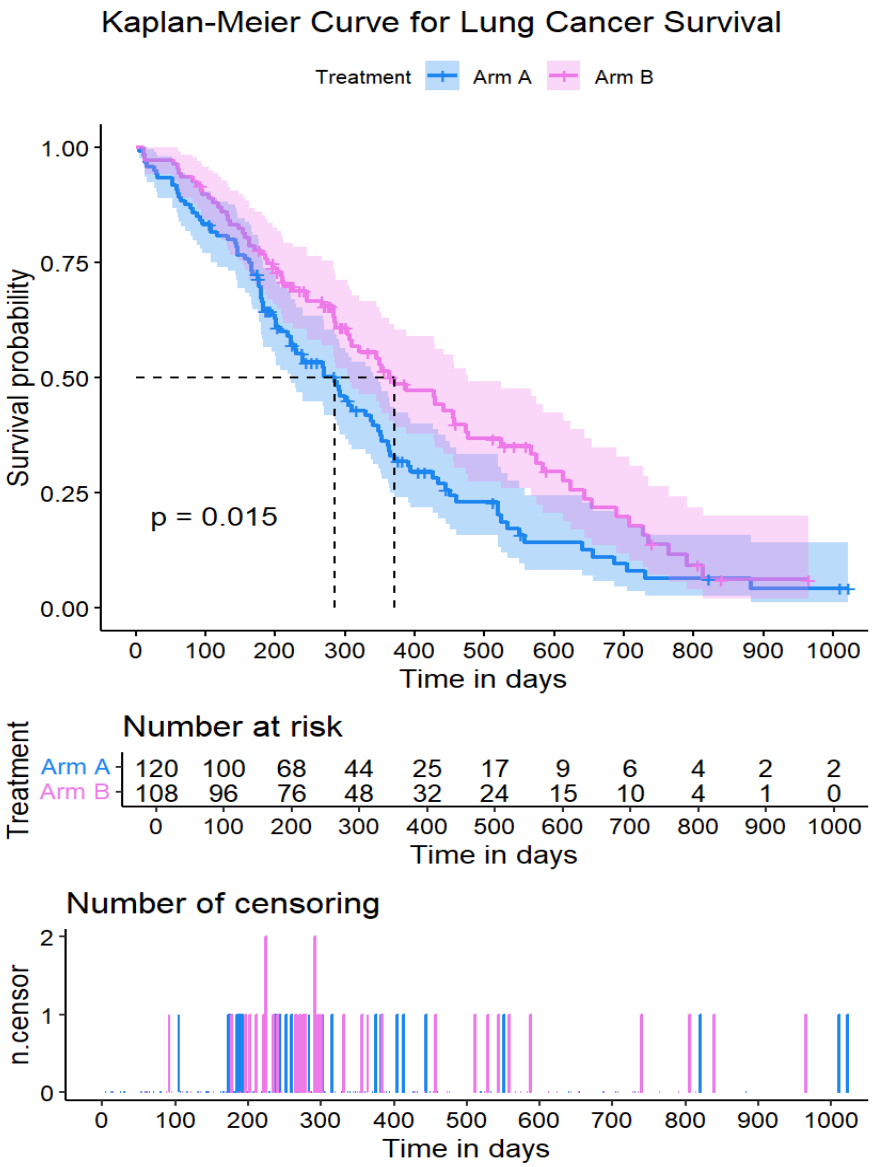

Рисунок 6. Графік виживання Каплан-Майера з додатковою інформацією.

На цьому графіку Каплан-Маєра горизонтальна вісь показує час у днях, а вертикальна вісь виживаємось, або частку людей які вижили. Про настанні події вказує вертикальний спад на кривій. Іноді спостереження завершуються перед тим, як подія відбудеться. Це називається цензурованими даними. На графіці цензуровані дані видаються вертикальними відрізками у точках, де спостереження закінчено, але подія не сталася.

По цьому графіку можна зробити висновки, що середнє виживання у пацієнтів, які приймали препарат В більша, приблизно 0.6 в момент часу 300

днів, ніж у пацієнтів, які приймали препарат А, приблизно 0,4 в момент часу 300 днів.

3.2.3 Графік кумулятивних подій.

Також доцільно побудувати графік кумулятивних подій, або кумулятивної захворюваності. В аналізі виживання кумулятивні події (cumulative events) є події, які відбуваються протягом певного періоду часу або на певних інтервалах часу. Ці події можуть бути смертю, виникненням ускладнень. Кумулятивні події вимірюються у міру того, як час минає. Кумулятивні події часто використовуються в контексті аналізу виживання для оцінки часу до настання певної події або групи подій. Графік кумулятивних подій (рис.7) часто будується разом із графіком виживання. Він є кривою, яка показує, як кількість подій накопичується в міру проходження часу або інтервалів часу [16].

```
> #cumulative events
> ggsurvplot(sfit, conf.int=TRUE, conf.int.style = "step",
+            xlab = "Time in days", surv.median.line = "hv",
+            ggtheme = theme_bw(), break.time.by = 100,
+            legend.labs=c("Arm A", "Arm B"), legend.title="Treatment",
+            palette=c("dodgerblue2", "orchid2"),
+            title="Kaplan-Meier Curve for Lung Cancer Survival",
+            fun = "event", xlim = c(0, 700))
```

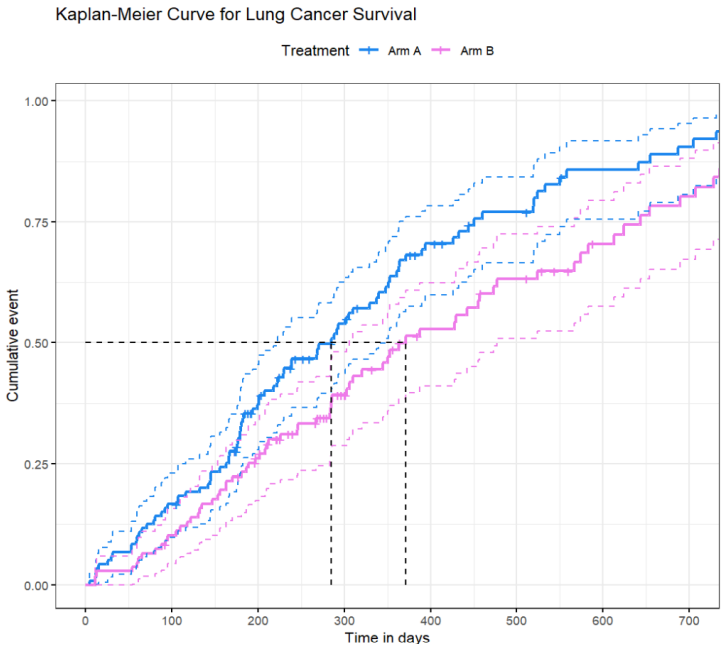

Рисунок 7.Графік кумулятивних подій.

3.2.4 Графік сукупної небезпеки.

Сукупна небезпека (Cumulative Hazard) - це концепція, що використовується в аналізі виживання для оцінки накопиченого ризику події до певного часу [13].

Сукупна небезпека є накопиченою функцію ризику (Hazard Function), яка є інтенсивністю виникнення події у кожен час від початку спостереження до часу t. Вона є сумою ризику (чи ймовірності) події на момент часу t. Формально, сукупна небезпека H(t) визначається як інтеграл від функції ризику h(t) від початку спостереження до часу t:

$$H(t) = \int_t^0 h(u)du \tag{6}$$

де:

- h(t) - функція ризику, яка є інтенсивністю виникнення події в момент часу t.
- H(t) - сукупна небезпека, що є накопиченим ризиком події на момент часу t.

Сукупна небезпека корисна для аналізу часу до виникнення події, як-от смерть, відмова устаткування тощо., вона дозволяє оцінити накопичений ризик на кожному кроці у процесі спостереження (рис.8). Це дозволяє порівнювати накопичені ризики між різними групами чи підгрупами та робити висновки про фактори, що впливають на ймовірність виникнення події [16].

```
> #cumulative hazard
> ggsurvplot(sfit, conf.int=TRUE,
+            xlab = "Time in days", surv.median.line = "hv",
+            ggtheme = theme_bw(), break.time.by = 200,
+            legend.labs=c("Arm A", "Arm B"), legend.title="Treatment",
+            palette=c("dodgerblue2", "orchid2"),
+            title="Kaplan-Meier Curve for Lung Cancer Survival",
+            fun = "cumhaz")
```

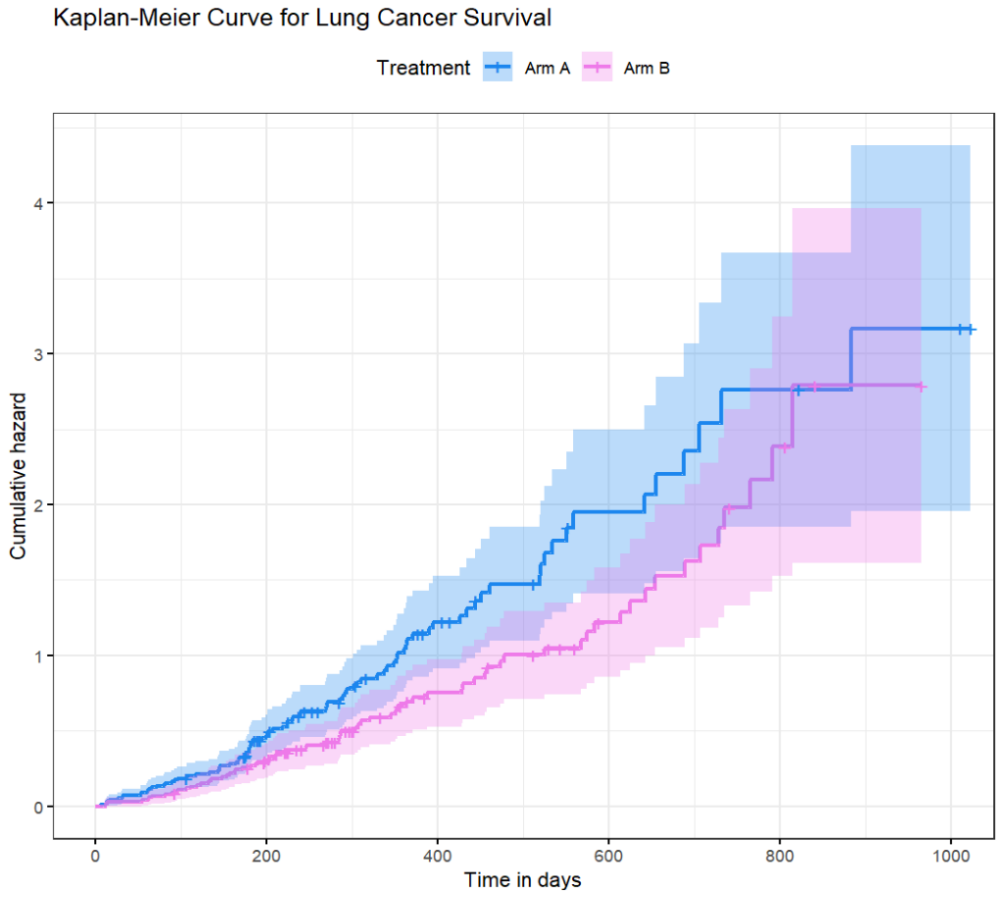

Рисунок 8. Графік сукупної небезпеки.

3.2.5 Тест лог-ранга (Log-Rank Test)

Тест Лог-Рангу (Log-Rank Test) є статистичним методом, який використовується в аналізі виживання для порівняння кривих виживання між двома або більше групами пацієнтів або подій. Він призначений для перевірки гіпотези про те, що немає відмінностей між функціями різних груп, що виживають.

Для кожної групи обчислюються очікувані та події, що спостерігаються. Очікувані події - це кількість подій, що очікувалося б, якби ймовірність події в кожному моменті часу була однаковою для всіх груп. Події, що спостерігаються - це фактична кількість подій, які відбулися в кожній групі. Обчислюється статистика тесту лог-рангу, яка порівнює спостережувану та

очікувану кількість подій у кожній групі. Статистика тесту підпорядковується хі-квадрат розподілу [17].

Функцію survdiff() можна використовувати для обчислення лог-ранг теста (рис.9):

```
> surv_diff <- survdiff(Surv(time, cnsr) ~ arm, data = lung_n)
> surv_diff
Call:
survdiff(formula = Surv(time, cnsr) ~ arm, data = lung_n)

        N Observed Expected (O-E)^2/E (O-E)^2/V
arm=1 120       93     77.6      3.08      5.89
arm=2 108       72     87.4      2.73      5.89

 Chisq= 5.9  on 1 degrees of freedom, p= 0.02
```

Рисунок 9. Функція survdiff()

Значення p-value дорівнює 0,02, це означає , що нульова гіпотеза може бути відкинута, тобто існує розбіжність між двома кривими виживання.

Тест логрангу є непараметричним тестом, що означає, що він не вимагає припущень про розподіл даних і підходить для аналізу невеликих вибірок. Він також чутливий до відмінностей у формі кривих виживання, що робить його методом, що широко використовується в аналізі виживання.

3.2.6 Криві виживання за декількома факторами.

За допомогою ggsurvplot() можна побудувати криві виживання для багатьох факторів одночасно (рис.10). Побудуємо криві виживання для параметру arm в залежності від параметрів sex та race.

```
require("survival")
fit2 <- survfit( Surv(time, cnsr) ~ arm + sex + race, data = lung_n )
# Plot survival curves by arm and facet by sex and race
ggsurv <- ggsurvplot(fit2, fun = "event", conf.int = TRUE,
                     ggtheme = theme_bw())
ggsurv$plot +theme_bw() +
  theme (legend.position = "right")+
  facet_grid(sex ~ race)
```

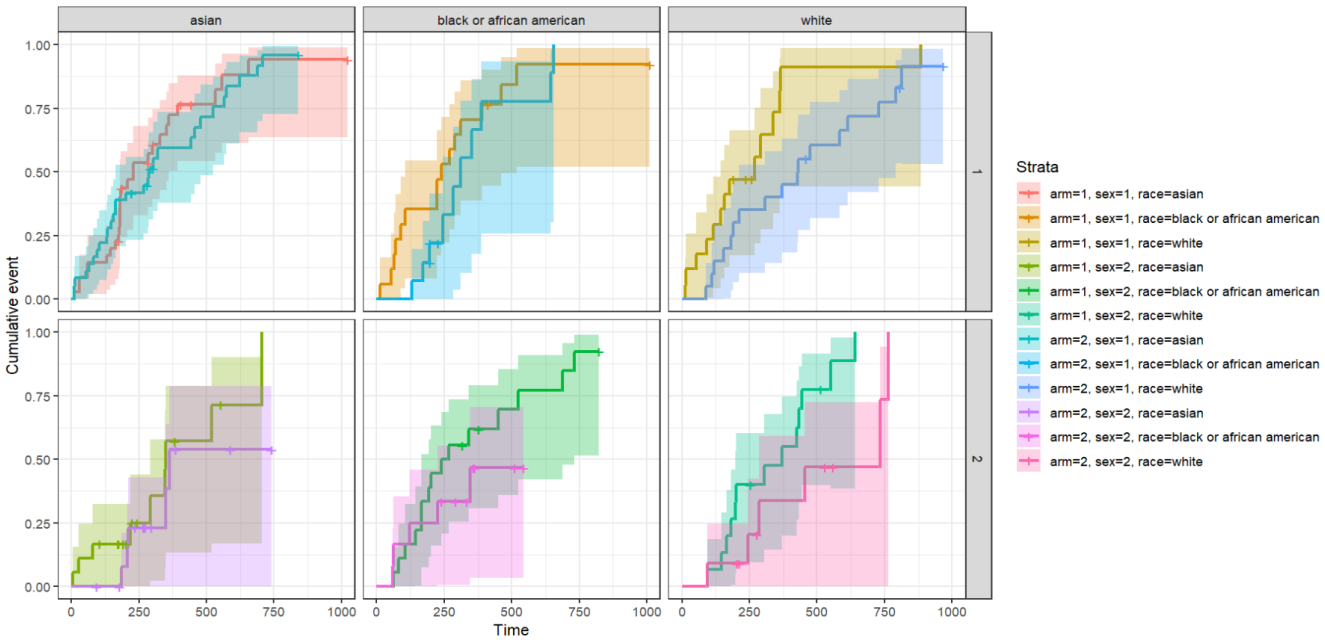

Рисунок 10. Криві виживання для фактору препарат, в залежності від статті та раси.

3.2.7 Однофакторний регресійний аналіз Кокса

Регресія Коксу, також відома як модель Коксу пропорційних ризиків, є статистичною моделлю, яка використовується в аналізі виживання для оцінки впливу одного фактору на виживання пацієнтів. У цій моделі передбачається, що вплив фактору на ризик події залишається з часом [18].

За допомогою мови R однофакторний аналіз Кокса можна обчислити (рис.11):

```
> res.cox <- coxph(Surv(time, cnsr) ~ arm, data = lung_n)
> res.cox
Call:
coxph(formula = Surv(time, cnsr) ~ arm, data = lung_n)

       coef exp(coef) se(coef)      z      p
arm -0.3812    0.6830   0.1579 -2.414 0.0158

Likelihood ratio test=5.89  on 1 df, p=0.01527
n= 228, number of events= 165
```

Рисунок 11. Регресія Коксу.

Функція *summary* () може створювати більш повний звіт (рис.12):

```
> summary(res.cox)
Call:
coxph(formula = Surv(time, cnsr) ~ arm, data = lung_n)

  n= 228, number of events= 165

       coef exp(coef) se(coef)      z Pr(>|z|)
arm -0.3812    0.6830   0.1579 -2.414   0.0158 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

    exp(coef) exp(-coef) lower .95 upper .95
arm     0.683      1.464    0.5012    0.9308

Concordance= 0.553  (se = 0.022 )
Likelihood ratio test= 5.89  on 1 df,   p=0.02
Wald test            = 5.83  on 1 df,   p=0.02
Score (logrank) test = 5.89  on 1 df,   p=0.02
```

Рисунок 12. Регресія Коксу (функція *summary* ())

Отримані результати можна інтерпретувати наступним чином:

1. Статистична значущість . Стовпець, «z», містить значення статистики Вальда. Він відповідає відношенню кожного коефіцієнта регресії до його стандартної помилки (z = coef/se(coef)). Статистика Вальда оцінює, чи бета (ββ) коефіцієнт даної змінної статистично значуще відрізняється від 0. З наведених вище даних можна зробити висновок, що змінна arm, тобто який препарат приймав пацієнт, має статистично значущі коефіцієнти.
2. Коефіцієнти регресії . Також при використанні моделі Кокса, треба звертати увагу на знак коефіцієнтів регресії (coef). Позитивний знак означає, що небезпека (ризик смерті) є вищою, а отже, прогноз гіршим для суб’єктів із вищими значеннями цієї змінної. Змінна arm (препарат):arm a (існуючий, контрольний препарат) = 1,):arm в (новітній, досліджуваний препарат) = 2. Результат для моделі Кокса дає коефіцієнт ризику (HR) для arm a відносно arm в. Бета-коефіцієнт для arm = -0,38говорить про те, що пацієнти які приймали новий препарат мають нижчий ризик смерті.

3. Коефіцієнти ризику . Експонентові коефіцієнти (exp(coef) = exp(-0,38) = 0,68), також відомі як співвідношення ризиків , дають величину ефекту коваріат. Тобто при прийомі нового препарату ризик поганої події зменшується на 32%.
4. Довірчі інтервали коефіцієнтів небезпеки . Підсумковий результат також дає верхні та нижні 95% довірчі інтервали для співвідношення ризиків (exp(coef)), нижня межа 95% = 0,50, верхня межа 95% = 0,93.
5. Глобальна статистична значущість моделі . Нарешті, вихідні дані дають p-значення для трьох альтернативних тестів загальної значущості моделі: тест співвідношення правдоподібності, тест Вальда та статистика логарифмічного рангу. Ці три методи асимптотично еквівалентні. Для достатньо великого N вони дадуть аналогічні результати. Для малих N вони можуть дещо відрізнятися.

Для застосування регресії Кокса одночасно для декількох коваріат (факторів): препарат, вік, стать, раса, ecog, зміна ваги, використовуємо наступний код (рис.13):

```
covariates <- c("arm", "age", "sex", "Race_Category", "ph_ecog", "wt_loss")
univ_formulas <- sapply(covariates,
                        function(x) as.formula(paste('Surv(time, cnsr)~', x)))

univ_models <- lapply( univ_formulas, function(x){coxph(x, data = lung_n)})
# Extract data
univ_results <- lapply(univ_models,
                       function(x){
                         x <- summary(x)
                         p.value<-signif(x$wald["pvalue"], digits=2)
                         wald.test<-signif(x$wald["test"], digits=2)
                         beta<-signif(x$coef[1], digits=2);#coeficient beta
                         HR <-signif(x$coef[2], digits=2);#exp(beta)
                         HR.confint.lower <- signif(x$conf.int[,"lower .95"], 2)
                         HR.confint.upper <- signif(x$conf.int[,"upper .95"],2)
                         HR <- paste0(HR, " (",
                                      HR.confint.lower, "-", HR.confint.upper, ")")
                         res<-c(beta, HR, wald.test, p.value)
                         names(res)<-c("beta", "HR (95% CI for HR)", "wald.test",
                                       "p.value")
                         return(res)
                         #return(exp(cbind(coef(x),confint(x))))
                       })
res <- t(as.data.frame(univ_results, check.names = FALSE))
as.data.frame(res)
```

Рисунок 13. Код регресії Коксу для декількох коваріат.

Як результат отримаємо (рис.14):

	beta	HR (95% CI for HR)	wald.test	p.value
arm	-0.38	0.68 (0.5-0.93)	5.8	0.016
age	0.022	1 (1-1)	5.5	0.019
sex	-0.35	0.7 (0.51-0.98)	4.4	0.037
Race_Category	-0.066	0.94 (0.78-1.1)	0.51	0.47
ph_ecog	0.52	1.7 (1.3-2.1)	22	3e-06
wt_loss	0.0013	1 (0.99-1)	0.05	0.83

Рисунок 14. Результат регресії Коксу для декількох коваріат.

З цього ми можемо зробити наступні важливі висновки:

1. Фактори препарат, вік, стать, ecog є статично значущими.
2. Фактори раса, зміна ваги не є статично значущими.
3. Фактори препарат, стать мають від'ємні бета-коефіцієнти, це означає, що жіноча стать та прийом новітнього препарату B веде до більшого виживання.
4. Фактори вік та ecog мають позитивні бета-коефіцієнти, тобто більший вік та значення ecog веде до зменшення виживання.

3.2.8 Багатофакторний регресійний аналіз Кокса

Багатофакторний регресійний аналіз Кокса, також відомий як мультифакторна модель Кокса, є статистичною моделлю, що використовується в аналізі виживання для вивчення впливу кількох незалежних змінних (факторів) на ризик виникнення події (наприклад, смерть) з часом [19].

Зараз, коли ми вже знаємо які фактори є статично значущими зробимо аналіз тільки для факторів препарат, вік, стать, ecog (рис.15)

```
> res.cox <- coxph(Surv(time, cnsr) ~ arm + age + sex + ph_ecog, data =  lung_n)
> summary(res.cox)
Call:
coxph(formula = Surv(time, cnsr) ~ arm + age + sex + ph_ecog,
    data = lung_n)

  n= 228, number of events= 165

             coef exp(coef) se(coef)      z Pr(>|z|)
arm      -0.39704   0.67231  0.16334 -2.431   0.0151 *
age       0.00843   1.00847  0.00954  0.884   0.3769
sex      -0.36925   0.69125  0.17047 -2.166   0.0303 *
ph_ecog   0.46898   1.59836  0.11197  4.188 2.81e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

         exp(coef) exp(-coef) lower .95 upper .95
arm         0.6723     1.4874    0.4881    0.9260
age         1.0085     0.9916    0.9898    1.0275
sex         0.6913     1.4466    0.4949    0.9655
ph_ecog     1.5984     0.6256    1.2834    1.9906

Concordance= 0.649  (se = 0.023 )
Likelihood ratio test= 32.89  on 4 df,   p=1e-06
Wald test            = 32.63  on 4 df,   p=1e-06
Score (logrank) test = 33.41  on 4 df,   p=1e-06
```

Рисунок 15. Аналіз Коксу для факторів препарат, вік, стать, ecog

За результатами ми можемо зробити наступні висновки:

1. P-значення для всіх тестів (вірогідність, Вальда та бал) є значущими, що вказує на значущість моделі. Ці тести оцінюють загальну нульову гіпотезу про те, що всі бета-версії (β) дорівнюють 0. Нульову гіпотезу обґрунтовано відхилено.
2. У багатофакторному аналізі Кокса фактори препарат, стать і ecog залишаються значущими ($p < 0,05$). Однак фактор вік не є значущим ($p = 0,37$, що більше ніж 0,05).
3. P-значення для типу препарату становіть 0,0151 із коефіцієнтом ризику HR = exp(coef) = 0,67, також P-значення для статі становить 0,0303 із HR = exp(coef) = 0,69, що вказує на зв’язок між типом препарату, статтю пацієнтів і зниженим ризиком смерті.
4. Подібним чином, p-значення для ph_ecog становить 2,81e-05, з коефіцієнтом ризику HR = 1,59, що вказує на тісний зв’язок між значенням ph.ecog і підвищеним ризиком смерті. Утримуючи інші

коваріати незмінними, більш високе значення ph.ecog пов’язане з поганою виживаністю.

5. Навпаки, p-значення для віку тепер становить p=0,37. Коефіцієнт ризику HR = exp(coef) = 1,01, з 95% довірчим інтервалом від 0,98 до 1,02. Тобто якщо інші фактори залишаються незмінними, додатковий рік віку спричиняє щоденну небезпеку смерті з коефіцієнтом exp(beta) = 1,008, або 0,8%, що не є значним внеском.

Висновок

За результатами отриманими в ході дослідження можна зробити наступні висновки:

- У момент часу 300днів ймовірність виживання становить приблизно 0,45 (або 45%) для arm a=1 і 0,75 (або 75%) для arm b=2.
- Середній час виживання для arm a=1 (контрольний препарат) становить 285 днів, на відміну від 371 днів для arm b=2 (досліджуваний препарат). Здається, досліджуваний препарат краще впливає на виживання ніж контрольний препарат.
- Фактори препарат, вік, стать, ecog є статистично значущими в дослідженні, тоді як фактори раса, зміна ваги не є статично значущими.
- Наявність жіночої статі та прийом досліджуваного препарату B позитивно впливає на виживання пацієнта, тоді як велике значення ecog навпаки негативно.
- Фактор вік, хоча і є статистично значущим не оказує великий вплив на виживання.
- Загалом, вибір між цими двома комплексами хіміотерапії повинен здійснюватися залежно від конкретних потреб і стану пацієнта. Рекомендується індивідуалізувати лікування, враховуючи показники та толерантність пацієнта до хіміотерапії.
- Для отримання більш точних результатів і визначення оптимального комплексу хіміотерапії для конкретних клінічних ситуацій можуть бути потрібні подальші дослідження, включаючи великі клінічні випробування.

Ці висновки слугують важливою основою для рекомендацій медичних фахівців та допомагають вибрати належний лікувальний підхід для пацієнтів у лікуванні певних патологій.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. https://dobro-clinic.com/ua/lechenie-raka/rak-legkikh/nemelkokletochnyj-rak-legkogo
2. https://radiologyassistant.nl/chest/lung-cancer-tnm-8th-edition
3. https://www.cancer.org/content/cancer/en/cancer/lung-cancer/detection-diagnosis-staging/signs-symptoms.html
4. https://www.webmd.com/lung-cancer/non-small-cell-lung-cancer#1
5. Aija Knuuttila / Настанова 00131. Рак легень // Настанови на засадах доказової медицини. Створені DUODECIM Medical Publications, Ltd. – 2017-03-21 https://guidelines.moz.gov.ua/documents/3032
6. E. Postmus, K.M. Kerr, M. Oudkerk, S. Senan, D.A. Waller, J. Vansteenkiste, C. Escriu and S. Peters / Early-Stage and Locally Advanced (non-metastatic) Non-Small-Cell Lung Cancer: ESMO Clinical Practice Guidelines // Ann Oncol (2017) 28 (suppl 4): iv1-iv21 https://doi.org/10.1093/annonc/mdx222
7. Kalbfleisch JD, Prentice RL. The Statistical Analysis of FailureTime Data. Hoboken, NJ: John Wiley & Sons, 2002;247-77.
8. Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the presence of competing risks: newrepresentations of old estimators. Stat Med. 1999;18(6):695-706.
9. Armitage P, Berry G, Matthews JN. 4th ed. Oxford (UK): Blackwell Science; 2002. Clinicaltrials. Statistical methods in medical research; C. 591.
10. Berwick V, Cheek L, Ball J. Statisticsreview 12: Survivalanalysis. CritCare. 2004; C. 389–394. [PMC free article] [PubMed]
11. Altman DG. London (UK): ChapmanandHall; 1992. Analysis of Survival times. In:Practical statistics for Medical research; C. 365–393.
12. Lee ET, Wang J. Statistical methods for survival data analysis. 2003; C. 476

13. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association. 1958; C. 457–481.
14. Klein JP, Moeschberger ML. Survival Analysis: Techniques for Censored and Truncated Data. 2nd ed. New York: Springer; 2010. [Google Scholar]
15. George B, Seals S, Aban I. Survival analysis and regression models. J Nucl Cardiol. 2014; 21:686–94.
16. Clark TG, Bradburn MJ, Love SB and Altman DG. Survival Analysis Part I: Basic concepts and first analyses. British Journal of Cancer (2003) 89, 232 – 238
17. Pocock S, Clayton TC, Altman DG (2002) Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. Lancet 359: 1686–1689.
18. Cox DR (1972). Regression models and life tables (with discussion). J R Statist Soc B 34: 187–220
19. MJ Bradburn, TG Clark, SB Love and DG Altman. Survival Analysis Part II: Multivariate data analysis – an introduction to concepts and methods. British Journal of Cancer (2003) 89, 431 – 436